An integrated bioinformatics analysis reveals divergent evolutionary pattern of oil biosynthesis in high- and low-oil plants

PLOS | ONE

RESEARCH ARTICLE

An Integrated Bioinformatics Analysis Reveals Divergent Evolutionary Pattern of Oil Biosynthesis in High- and Low-Oil Plants

Li Zhang[1], Shi-Bo Wang[1,2], Qi-Gang Li[3], Jian Song[1], Yu-Qi Hao[1], Ling Zhou[1,4], Huan-Quan Zheng[5], Jim M. Dunwell[6], Yuan-Ming Zhang[1,2]*

1 State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, People's Republic of China, **2** Statistical Genomics Laboratory, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, People's Republic of China, **3** State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, People's Republic of China, **4** Institute of Biotechnology, Jiangsu Academy of Agricultural Science, Nanjing 210014, People's Republic of China, **5** Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada, **6** School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AS, United Kingdom

* soyzhang@njau.edu.cn; soyzhang@mail.hzau.edu.cn

Citation: Zhang L, Wang S-B, Li Q-G, Song J, Hao Y-Q, Zhou L, et al. (2016) An Integrated Bioinformatics Analysis Reveals Divergent Evolutionary Pattern of Oil Biosynthesis in High- and Low-Oil Plants. PLoS ONE 11(5): e0154882. doi:10.1371/journal.pone.0154882

Editor: Thierry Chardot, INRA, FRANCE

Received: November 18, 2015

Accepted: April 20, 2016

Published: May 9, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the National Natural Science Foundation of China (grant 31571268), and Huazhong Agricultural University Scientific and Technological Self-innovation Foundation (Program No. 2014RC020).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Seed oils provide a renewable source of food, biofuel and industrial raw materials that is important for humans. Although many genes and pathways for acyl-lipid metabolism have been identified, little is known about whether there is a specific mechanism for high-oil content in high-oil plants. Based on the distinct differences in seed oil content between four high-oil dicots (20~50%) and three low-oil grasses (<3%), comparative genome, transcriptome and differential expression analyses were used to investigate this mechanism. Among 4,051 dicot-specific soybean genes identified from 252,443 genes in the seven species, 54 genes were shown to directly participate in acyl-lipid metabolism, and 93 genes were found to be associated with acyl-lipid metabolism. Among the 93 dicot-specific genes, 42 and 27 genes, including CBM20-like *SBDs* and *GPT2*, participate in carbohydrate degradation and transport, respectively. 40 genes highly up-regulated during seed oil rapid accumulation period are mainly involved in initial fatty acid synthesis, triacylglyceride assembly and oil-body formation, for example, *ACCase*, *PP*, *DGAT1*, *PDAT1*, *OLEs* and *STEROs*, which were also found to be differentially expressed between high- and low-oil soybean accessions. Phylogenetic analysis revealed distinct differences of oleosin in patterns of gene duplication and loss between high-oil dicots and low-oil grasses. In addition, seed-specific *GmGRF5*, *ABI5* and *GmTZF4* were predicted to be candidate regulators in seed oil accumulation. This study facilitates future research on lipid biosynthesis and potential genetic improvement of seed oil content.

Introduction

The angiosperms are the most diverse group of land plants with the number of species in the range of 250,000 to 400,000 [1]. They have dramatic differences not only in organ morphology (leaves, flowers, seeds, roots, and vascular tissues) but also in the chemical composition of the seed [2]. Interestingly, oil plants, such as peanut, sesame and soybean, are generally rich in oil of their seeds, while most cereals like rice, wheat and sorghum specifically accumulate starch and have a relatively low oil content in the seed. Since the formation of seed oil and starch is both dependent on the supply of photosynthetic carbon [3, 4], it is likely that there are divergent mechanisms in the evolution of these high- and low-oil plants to regulate the partitioning of carbon between oil and other storage products. Previous studies have demonstrated that increasing the carbon flow to lipid biosynthesis can significantly increase the seed oils [4–6]. However, the mechanism by which more carbohydrates flow to *de novo* fatty acid (FA) synthesis in high-oil content plants is unclear.

Seed oil is not only the major source of carbon and energy for germination and seedling growth but also provides humans with renewable sources of food, biofuel and industrial raw materials. Up to now acyl-lipid metabolism in Arabidopsis has been well studied and more than 600 genes have been predicted to encode the enzymes and regulatory factors associated with this process [7, 8]. Among these predicted genes, some have been shown to be associated with changes in seed oil accumulation [6]. The over-expression of many individual key enzymes altered seed oil content in various plants. For example, ACCase in potato [9], *Brassica napus* [10] and *Escherichia coli* [11]; glycerol-3-phosphate dehydrogenase (GPDH) in *B. napus* [12]; glycerol-3-phosphate acyltransferase (GPAT) in *Arabidopsis* [13]; 2-lysophosphatidic acid acyltransferase (LPAAT) in *Arabidopsis* [14]; and acyl-CoA: diacylglycerol acytransferase (DGAT) in *Arabidopsis* [15, 16], *Glycine max* [17], *B. napus* [18] and maize [19]. In addition to these key enzymes that participate in lipid synthesis, the expression of transcription factors (TFs) directly or indirectly regulating genes involved in carbohydrate and lipid metabolism can also evidently change seed oil content; such TFs include *WRINKLED1* (*WRI1*) [20], *LEAFY COTYLEDON1* (*LEC1*) [21, 22], *LEAFY COTYLEDON2* (*LEC2*) [23], *FUSCA3* (*FUS3*) [24], *GmbZIP123* [25], *GmMYB73* [26] and *ABSCISIC ACID INSENSITIVE3* (*ABI3*) [27, 28]. However, each of the above studies only focused on a single enzyme or TF involved in lipid metabolism. In reality, seed oil content is affected by multiple genes [29, 30] or multiple reactions [31]. For example, specific combination of expression of *WRI1*, *DGAT* and triacylglycerol lipase *SUGAR-DEPENDENT1* resulted in a higher percentage seed oil content than that obtained by manipulation of each gene individually [32]. More importantly, the seed oil content is influenced by multiple metabolic pathways, such as sucrose catabolism, glycolysis, pentose phosphate pathway, and related pathways which rely on the supply of carbon [33–38]. Therefore, it is necessary to consider the involvement of multiple genes and the interaction of multiple related pathways in order to understand the mechanism of high-oil content in high-oil plants. The recent sequencing of many plant genomes has provided an opportunity to investigate this mechanism using comparative genome and transcriptome analyses.

Analysis of genome-wide differential gene expression between developmental stages, and between sub-species (like 输入文字或网址，即可翻译 cultivated and wild forms) could provide insights into biological pathways and molecular mechanisms that regulate seed development and nutrient accumulation. Seed development is also an important part of the reproductive (R) process in flowering plants. In soybean, this reproductive process is divided into eight stages from flowering (R1) to full maturity (R8) [39]. During the R4 to R7 stages, importantly, seeds grow rapidly, accumulating nutrients, lipids and storage proteins.

In this study, all the genes from three low-oil grasses (*Sorghum bicolor*, *Setaria italica*, and *Oryza sativa*) and four high-oil dicots (*Glycine max*, *Gossypium raimondii*, *Ricinus communis*, and *Arabidopsis thaliana*) were clustered in order to obtain a list of high-oil dicot-specific genes. A gene ontology (GO) enrichment analysis and a pathway level co-expression (PLC) network analysis were then conducted to identify genes or TFs that are likely to be associated with oil accumulation. Analyses of gene expression during the various stages of seed development in soybean and of RNA sequencing (RNA-seq) differential expression between high- and low-oil content soybean accessions were performed to identify differentially expressed genes (DEGs) in the core pathways associated with the deposition of seed oil. These results were used to further investigate how evolutionary divergence contributes to differences in seed oil content between the two kinds of plants and to discover new genes associated with the seed oil differences.

Results

Identification and GO enrichment analysis of dicot-specific genes

In this study, OrthoMCL was applied to construct potential orthologous groups (OGs) of proteins across four high-oil dicots (seed oil content: 20–50%) and three low-oil grasses (<3%) (S1 Table), because it can group both orthologs and paralogs over multiple eukaryotic taxa by using a Markov Cluster algorithm (MCL) [40]. As a result, all the 252,443 genes from the above seven species were clustered into 29,095 OGs (S1 Dataset). Among these gene families, only 1,534 (5.27%) OGs appear to be specific to the high-oil dicot lineage and are defined as high-oil dicot-specific clusters since all the genes in these OGs come from all the four rosid species but not from any grasses (S1 Dataset). Note that OrthoMCL clusters proteins based on overall conservation but not on individual protein domains. Thus, high-oil dicot-specific OGs in this study contain families specific in high-oil dicots or families with much low similarities between dicots and grasses. Among the 1,534 high-oil dicot-specific clusters, 4,051, 2,758, 1,731 and 2,152 genes were found in *G. max*, *G. raimondii*, *R. communis*, and *A. thaliana*, respectively.

To understand the functions of these high-oil dicot-specific genes, a GO enrichment analysis was conducted for 4,051 soybean genes compared with all the annotated genes. 392 GO terms for biological processes, molecular functions and cellular components were identified; and these were distributed in 77 GO slim terms (S2 Dataset).

Among the 43 GO slims for biological processes, some were involved in metabolic pathways (S2 Dataset), such as biosynthetic process, carbohydrate metabolic process, catabolic process, generation of precursor metabolites and energy, lipid metabolic process, metabolic process, protein metabolic process and secondary metabolic process. In addition, slim GO:0006810 (transport) was involved in transport of many intermediates of carbohydrate degradation and glycolytic pathways, which includes triose phosphate transmembrane transport, phosphoenolpyruvate transport, acylglycerol transport, glucose-6-phosphate transport, phosphoglycerate transport, hexose phosphate transport, regulation of intracellular transport and triose phosphate transport.

Expression patterns of dicot-specific genes during seed development

The transcriptomic data for soybean seeds at seven stages of development [41] were used to analyze the expression patterns of 4,051 dicot-specific genes. 3,155 (77.88%) genes were expressed in developing seeds across two biological replicates. Among these 3,155 genes, 3,150 (99.84%) were grouped into eight clusters based on Pearson's correlation coefficients,

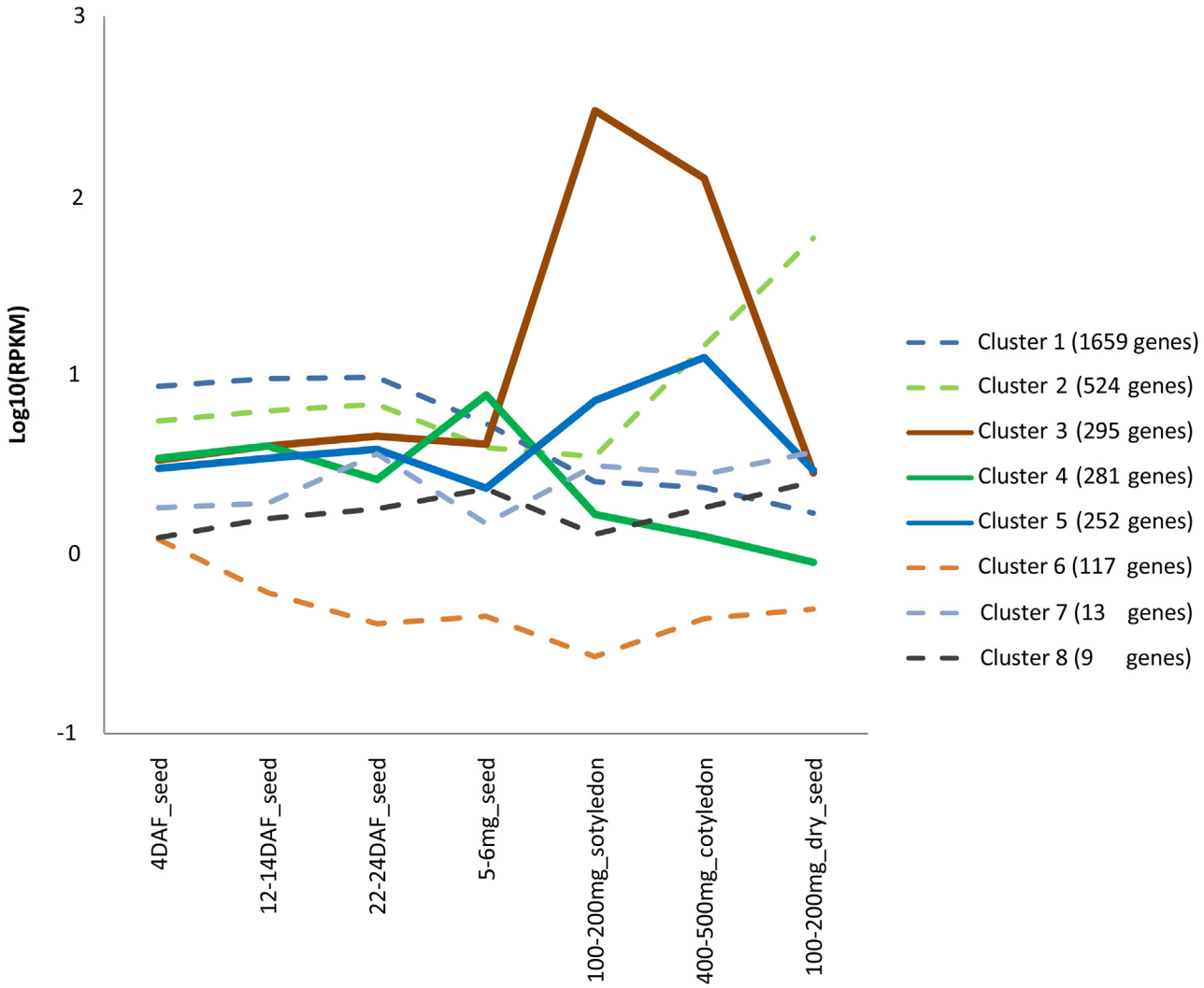

Fig 1. Average expression profile (*y* axis) for each cluster of dicot-specific genes at seven stages of soybean seed development (*x* axis). RPKM: reads per kilobase of gene model per million mapped reads; DAF: days after flowering. The dataset was downloaded from http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42871.

doi:10.1371/journal.pone.0154882.g001

implemented by MCL [42]. All the eight clusters are shown in Fig 1 and the genes in each cluster are listed in S1 Dataset.

Identification of genes related to seed oil content in soybean

Identification of high-oil dicot-specific genes directly participating in acyl-lipid metabolism. Compared with acyl-lipid metabolism genes in *Arabidopsis* [7, 8], 1,123 orthologous genes were identified in soybean (S1 Dataset). Among these 1,123 genes, 54 were high-oil dicot-specific genes and distributed in the above eight clusters (S2 Table). In particular, almost all the genes coding key enzymes of fatty acid synthesis like Biotin Carboxylase (BC), Biotin Carboxyl Carrier Protein 2 (BCCP2), Carboxyltransferase alpha subunit (α-CT), WRI1, and FUS3 were in cluster 3.

Dicot-specific lipid-metabolism-related genes predicted from PLC analysis. Genes in the clusters 3 to 5 show highest expression levels at one of the stages from 5–6 mg to 400–500

mg, which tend to occur between R4 and R7. During these stages, lipids are rapidly accumulated in the seed [39]. To ensure prediction accuracy, these genes in clusters 3 to 5 were selected for further analysis. This is because that the genes in cluster 2 show highest expression at the dry whole seed stage and the ones in the remaining clusters are down-regulated during seed development. In the clusters 3 to 5, there were 828 genes (S1 Dataset), among which 23 genes have been determined to participate in the above lipid synthesis (S2 Table). GO enrichment analysis for the remaining 805 genes assigned 207 (25.71%) genes to the biological processes that may be associated with carbohydrate and lipid metabolic pathways (S3 Dataset and S3 Table), such as carbohydrate metabolic process, lipid metabolic process, transport, signal transduction, and catabolic process.

To further identify and prioritize novel candidate acyl-lipid metabolism members in these three clusters, we conducted a pathway-level coexpression (PLC) network analysis [43–46] between the above 207 genes and 1,123 genes in acyl-lipid metabolism. The results show that 93 dicot-specific genes are candidate genes involved in acyl-lipid metabolism pathways (Table 1, S4 Table and S5 Dataset).

Expression patterns of genes encoding core lipid synthetic enzymes. In soybean, the core pathways for accumulating seed oil operate through FA synthesis and the export of FAs from the plastid followed by triacylglycerol synthesis, and oil body formation, which included 156 genes (S1 Dataset). Among these genes, 113 (72.44%) had higher expression level than the average in at least one of the stages from 5–6 mg to 400–500 mg (S4 Dataset), indicating that most key genes for lipid synthesis were up-regulated during oil accumulation in *G. max*. This phenomenon was similar to that in *Arabidopsis* [47]. If the criterion used was twice average level, 40 (25.6%) genes were identified (S4 Dataset). These 40 genes were largely distributed in the initiation of FA synthesis, triacylglycerol synthesis, and oil-body formation (Fig 2).

Fourteen of the above 40 genes were observed in FA synthesis, and 10 of the 14 genes were involved in the initiation of FA synthesis. Among the 10 genes, five encoded two subunits of the plastidial pyruvate dehydrogenase complex (PDHC) that promotes pyruvate decarboxylation to acetyl-CoA and is the key enzyme linking carbohydrate metabolism to FA synthesis [4]; and the others encoded two subunits of ACCase. Of these 40 genes, two genes, those encoding ACP4 and LACS9, were involved in FA transportation. Eight of the above 40 genes were involved in TAG assembly; this group included 2, 3, 2, and 1 genes respectively coding phosphatidate phosphatase (PP), DGAT1, phospholipid: diacylglycerol acyltransferase (PDAT1), and PDAT2; these enzymes catalyze the consecutive steps after the second acylation of glycerol-3-phosphate [15, 19, 48]. Three and six of the above 40 genes, respectively encoding steroleosins (STEROs) and oleosins (OLEs), were observed in oil-body formation, and they were up-regulated, mostly by some hundredfold, during stages from 5–6 mg to 400–500 mg in seeds (S4 Dataset). The remaining genes in this group of 40 genes encoded the TFs WRI1, VAL1, FUS3, and ABI3. Kim et al. [49] divided Arabidopsis oleosin genes into three groups on the basis of their tissue-specific expression. In this study, all the genes in the S type (genes expressed only in maturing seed (siliques)) and SM type (genes expressed in both maturing seeds and florets (microspores)) were used to conduct a phylogenetic analysis (Fig 3). The results showed distinct differences of OLEs in patterns of gene duplication and loss between high-oil dicots and low-oil grasses. Genes coding oleosins were preferentially retained in high-oil dicots. In two angiosperm groups, there was only one copy in each grass species, while there were many duplicates in dicots. Remarkably, two high-oil dicot-specific groups were identified, which were resulted from gene loss in grasses after whole genome duplication of ancestor angiosperm. In dicot-specific group 1, *Glyma04g08220* and *Glyma06g08290* were up-regulated by a thousand-fold during stages from 5–6 mg to 400–500 mg (S4 Dataset) and may play an important role in soybean seed oil accumulation. In dicot-specific group 2, *At3G01570*,

Table 1. Dicot-specific genes associated with acyl-lipid metabolism.

GO slims	Dicot-specific genes	Pathway annotation	Pfam annotation	Enzymes/Proteins	Reference
GO:0005975; Carbohydrate metabolic process	*Glyma0165s00200*, *Glyma01g28520*, *Glyma03g08860*		PF00686	*CBM20-like* starch binding domian	Marchler-Bauer et al. [67]; Southall et al. [70]; Rodriguez-Sanoja et al. [69]; Christiansen et al. [68]
	Glyma09g04330, *Glyma15g15360*	PWY-6902; chitin degradation II	PF00182	chitinase	
	Glyma01g33440, *Glyma03g03400*, *Glyma03g03460*, *Glyma06g47690*, *Glyma06g47690*, *Glyma12g00700*	PWY-1081; homogalacturonan degradation	PF01095, PF04043	pectinesterase	
	Glyma04g02660, ***Glyma06g02690***, *Glyma14g40400*, ***Glyma17g37750***, ***Glyma17g37760***		PF02704	Gibberellin-regulated family protein	Aubert et al. [97]; Chen et al. [98]
	Glyma01g37150, ***Glyma11g08120***, *Glyma15g08960*, *Glyma05g36680*, *Glyma14g40110*, ***Glyma02g34870***, *Glyma03g36250*, ***Glyma10g10530***, ***Glyma19g38900***, *Glyma02g36150*, *Glyma10g08740*, *Glyma08g16880*, *Glyma09g01520*, *Glyma17g36200*, *Glyma14g08970*, *Glyma10g16100*, *Glyma07g05970*, *Glyma04g40630*, *Glyma06g14160*, *Glyma05g06410*, *Glyma19g07830*, *Glyma18g00560*, *Glyma02g47500*, ***Glyma14g01260***, *Glyma17g09260*		S4 Table		
GO:0006810; Transport	*Glyma07g38830*, *Glyma13g27680*, ***Glyma15g11270***		PF03151, PF00892	*GPT2*	Kammerer et al. [71]; Niewiadomski et al. [74]; Andriotis et al. [73]; Kunz et al. [75]; Bourgis et al. [77]
	Glyma18g08740		PF03151	triosephosphate translocator subfamily protein	
	Glyma12g13070, ***Glyma08g02510***, *Glyma03g05880*, *Glyma15g17310*, *Glyma06g40980*, *Glyma07g12460*, *Glyma09g06260*, *Glyma12g34020*, *Glyma18g14810*, *Glyma09g04870*, ***Glyma15g15990***, *Glyma04g06230*, *Glyma03g33150*, *Glyma17g09640*, ***Glyma05g37050***, *Glyma09g08410*, *Glyma12g33880*, *Glyma19g34820*, *Glyma03g33660*, *Glyma06g22240*, *Glyma14g33680*, *Glyma20g23120*, *Glyma17g09260*		S4 Table		
GO:0006629; Lipid metabolic process	*Glyma02g45680*, *Glyma14g03130*		PF00067	cytochrome P450, family 718	Pinot and Beisson [99]; Sun et al. [100]
	Glyma18g45250	PWY-2761; glyceollin biosynthesis I	PF01370	pterocarpin synthase	Kim et al. [101]
	Glyma02g46220, *Glyma14g02510*		S4 Table		
GO:0007165; Signal transduction	*Glyma09g07090*, ***Glyma15g18380***, *Glyma17g06290*		PF00320	GATA type zinc finger transcription factor family protein	Bi et al. [96]; Velmurugan et al. [95]
	Glyma09g39570, *Glyma10g12130*	PWY-5035; gibberellin biosynthesis III	PF03171	gibberellin 3β-dioxygenase	Chen et al. [98]

(*Continued*)

Table 1. (*Continued*)

GO slims	Dicot-specific genes	Pathway annotation	Pfam annotation	Enzymes/Proteins	Reference
	Glyma12g32910, ***Glyma13g39340***, *Glyma02g47500*, *Glyma14g01260*, *Glyma13g29070*, *Glyma15g09980*		S4 Table		
GO:0019538; Protein metabolic process	*Glyma05g23620*, *Glyma17g16690*, *Glyma01g03580*, *Glyma06g10110*, *Glyma08g39290*, *Glyma17g12200*, *Glyma18g19720*		S4 Table		

High-oil dicot-specific genes were predicted by PLC analysis. 17 genes with bold type were also differentially expressed in the high- and low-oil soybean seeds at the 0.01 significant level. All the information containing both annotation and P-value in differential expression analysis for each gene is given in S4 Table.

doi:10.1371/journal.pone.0154882.t001

At3G27660 and *At5G40420* in Arabidopsis were shown to be S type genes [49]. Meanwhile, *Glyma05g08880* and *Glyma19g00400* in this group were highly up-regulated in hundred-fold in the dry whole seed (S4 Dataset), indicating that this dicot-specific group may perform some specific functions in seed development. Furthermore, highly up-regulated and abundant OLEs probably play major roles in oil accumulation and/or oil body development, and are associated with high-oil content in seeds [50, 51].

Interestingly, *OLEs*, *STEROs* and *ABI3* genes are all significantly over-expressed and co-expressed with each other (S1 Fig and S4 Dataset). Since *OLEs* were regulated by *ABI3* [28], *ABI3* might also up-regulate the expression of *STEROs*. To improve seed oil content in molecular breeding, it may be beneficial to increase the expression of *OLEs*, *STEROs* or *ABI3* genes.

Identification of candidate TF genes related to seed oil accumulation. Several important transcription factors (TFs) have been found to participate in the regulation of seed oil accumulation. These TFs are distributed in the TF families like AP2, basic leucine zipper (bZIP), B3, NF-YB and Dof [21, 22, 24–27, 52, 53]. In soybean, there were 3,714 genes coding TFs [54], and 19 of these were found to be known lipid-related TFs (S5 Dataset). To identify additional potential TFs related to seed oil accumulation, we conducted a PLC network analysis between all the TFs in soybean and the above 156 genes coding core lipid synthetic enzymes. As a result, 327 TFs were found to be co-expressed with at least two genes coding core lipid synthetic enzymes (P-value<1e-4) (S5 Dataset), and were classified into 42 gene families. Among the 327 TFs, 67 were distributed in high-oil dicot-specific clusters (S5 Dataset), and 70 had expression levels greater than twice the average level in at least one of stages from 5–6 mg to 400–500 mg. These 70 TFs were compared with the 69 seed-specific TFs described by Song et al. [25]. Seven genes were found to be seed-specific, respectively coding tandem CCCH zinc finger protein 4 (*GmTZF4*; *Glyma12g13300*, *Glyma06g44440*, *Glyma12g33320*), growth-regulating factor 5 (*GmGRF5*; *Glyma07g04290*), ABI5 (*Glyma10g08370*), ABI3 (*Glyma18g38490*) and a sequence-specific DNA binding TF (*Glyma03g34730*) (Table 2). Since these seven seed-specific TFs were not only highly up-regulated during the rapid oil accumulation phase of seed development but also co-expressed with key enzymes of lipid synthesis genes, we deduced that they possibly play roles in regulating seed oil accumulation. It should be noted that ABI3 has been confirmed to be a regulator in oleosin gene expression [17, 28].

We investigated co-expression networks of the above 19 known lipid-related TFs with 156 genes coding core lipid synthetic enzymes. As a result, 14 lipid-related TFs were co-expressed with more than 10 lipid synthesis genes (P-value<0.05) (S5 Dataset), such as two genes (*Glyma15g34770* and *Glyma08g24420*) coding WRI1 were found to be co-expressed with 19 lipid

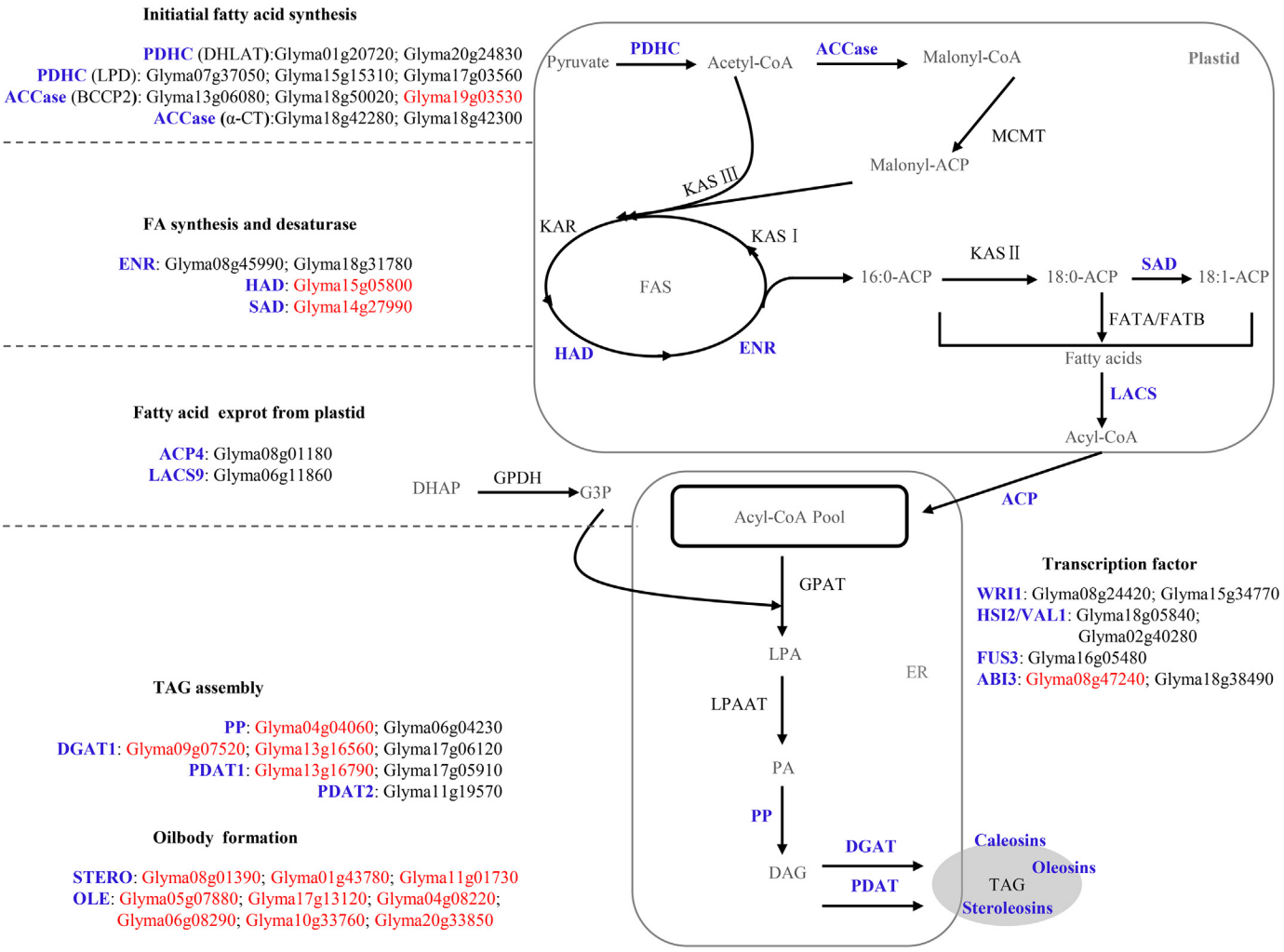

Fig 2. The distribution of 40 highly up-regulated genes and the respective DEGs in the framework of acyl-lipid metabolism. Enzymes with blue color are highly up-regulated and 40 genes coding them are listed. Among these 40 highly up-regulated genes, DEGs between the high- and low-oil soybean accessions are highlighted with red color. ABI, ABSCISIC ACID INSENSITIVE; ACP, acyl carrier protein; BCCP, biotin carboxyl carrier protein; DGAT, acyl-CoA: diacylglycerol acytransferase; DHAP, Dihydroxyacetone phosphate; DHLAT, dihydrolipoamide acetyltransferase; ENR, enoyl-ACP reductase; ER, endoplasmic reticulum; FAS, fatty acid synthase; FATA(B), fatty acylthioesteraseA(B); FUS, FUSCA; G3P, glycerol-3-phosphate; GPAT, glycerol-3-phosphate acyltransferase; GPDH, glycerol-3-phosphate dehydrogenase; HAD, hydroxyacyl-ACP dehydrase; KAR, ketoacyl-ACP reductase; KAS, ketoacyl-ACP synthase; LACS, long-chainacyl-CoA synthetase; LPD, dihydrolipoamide dehydrogenase; MCMT, malonyl-CoA:ACP malonyltransferase; OLE, oleosins; PDAT, phospholipid:diacylglycerolacyl transferase; PDH, pyruvatede hydrogenase; PP, phosphatidate phosphatase; SAD, stearoyl-ACP desaturase; STERO, steroleosin; TAG, triacylglycerol; WRI, WRINKLED.

doi:10.1371/journal.pone.0154882.g002

synthesis genes (Table 3). Among the 19 lipid synthesis genes, genes coding BCCP2, BC, ENR and α-PDH have been verified to be regulated by *WRI1* in *Arabidopsis* [53, 20, 23], and 18 genes have an AW-box in their promoter (S2 Fig) that has been shown to be a direct target of WRI1 in *Arabidopsis* [53]. Note that *WRI1* has a high co-expression relationship with *Glyma06g11860* (*LACS9*; r = 0.9801, P-value = 2.32e-5), and *LACS9* is found to impact the biosynthesis of seed storage lipids in *Arabidopsis* [55], and is considered as the major LACS isoform involved in plastidial FA export for TAG formation [56]. We surmised that *WRI1* up-regulated *LACS9* and this regulation could lead to an increased export of FAs to the endoplasmic reticulum (ER), and a subsequent increase in the rate of FA synthesis and triacylglycerol synthesis.

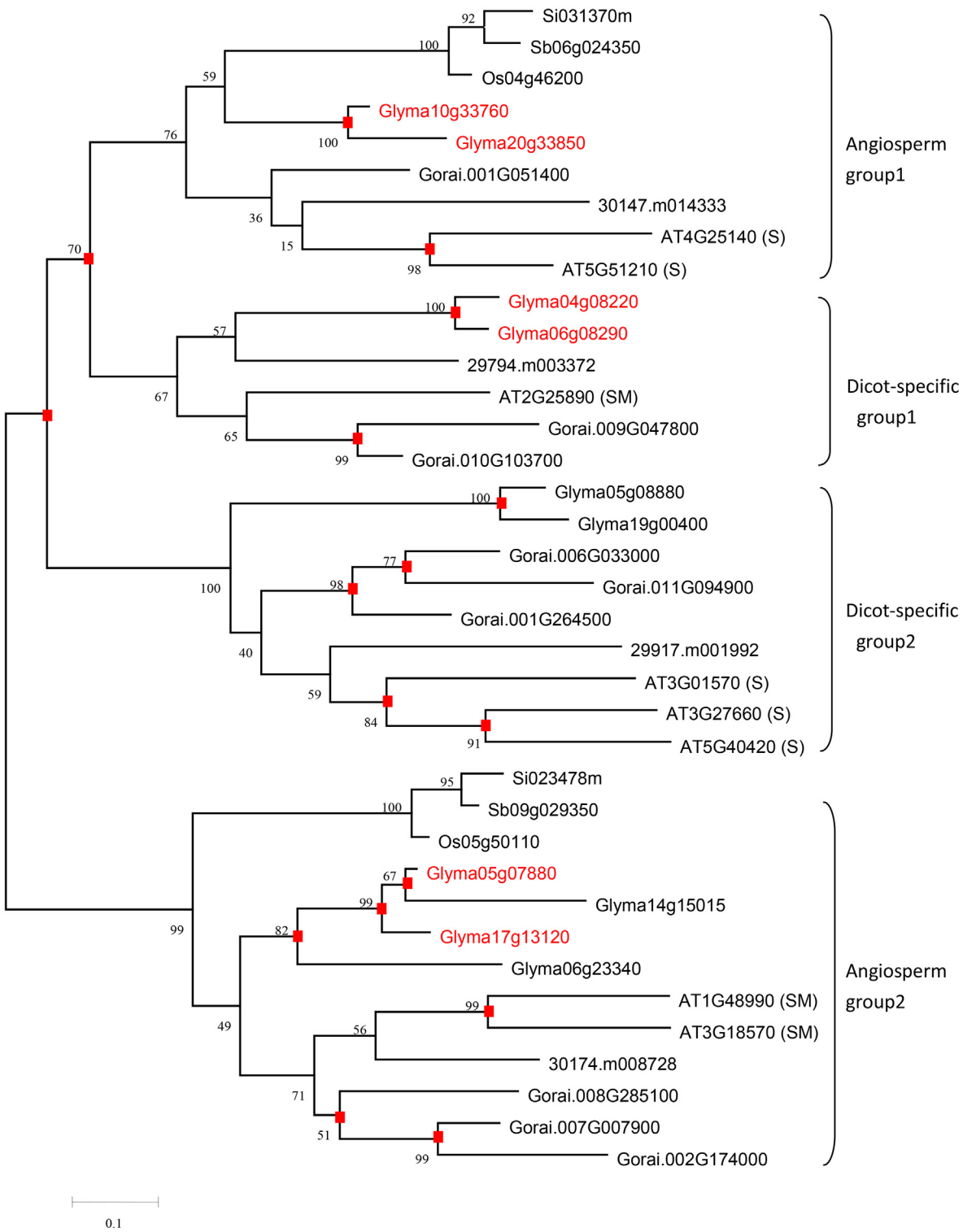

Fig 3. Phylogenetic tree of oil body protein OLE. The phylogenetic tree was constructed using Neighbor-Joining method. The numbers on the branches represent the bootstrap support, and square boxes indicate duplication events. S indicates genes expressed only in maturing seed and SM in both maturing seeds and florets. The highly up-regulated genes in Fig 2 are highlighted in red color.

doi:10.1371/journal.pone.0154882.g003

Table 2. Pathway level co-expression analysis between seven seed-specific transcript factors and genes in the key pathways of lipid synthesis.

Transcript factor	Subfamily	P-value in differential expression analysis		Pathway-level co-expression analysis				Arabidopsis homolog of soybean transcript factor
		Group H-L1	Group H-L2	Gene	Abbreviation	r	P-value	
Glyma12g13300 (*GmTZF4*)	C3H	3.73e-34	3.10e-09	*Glyma18g05840*	*VAL*	0.9728	5.87e-05	*At1g03790* (*AtTZF4*; *SOM*)
				Glyma04g08220	*OLE*	0.9930	1.05e-06	
				Glyma06g08290	*OLE*	0.9891	3.87e-06	
				Glyma09g07520	*DGAT1*	0.9683	9.26e-05	
				Glyma08g01390	*STERO*	0.9997	8.20e-11	
				Glyma07g32850	*SAD*	0.9898	3.16e-06	
Glyma06g44440 (*GmTZF4*)	C3H	6.21e-31	1.29e-09	*Glyma19g03530*	*BCCP2*	0.9757	4.22e-05	*At1g03790* (*AtTZF4*; *SOM*)
				Glyma04g08220	*OLE*	0.9855	9.10e-06	
				Glyma09g07520	*DGAT1*	0.9929	1.08e-06	
				Glyma11g19570	*PDAT2*	0.9717	6.62e-05	
				Glyma08g01390	*STERO*	0.9871	6.39e-06	
				Glyma07g32850	*SAD*	0.9768	3.68e-05	
Glyma12g33320 (*GmTZF4*)	C3H	8.79e-37	1.14e-07	*Glyma03g29600*	*LPAAT*	-0.9759	4.08e-05	*At1g03790* (*AtTZF4*; *SOM*)
				Glyma19g32420	*LPAAT*	-0.9705	7.43e-05	
				Glyma18g38490	*ABI3*	0.9904	2.66e-06	
Glyma07g04290 (*GmGRF5*)	GRF	2.20e-13	3.15e-08	*Glyma09g07520*	*DGAT1*	0.9915	1.84e-06	*At3g13960* (*AtGRF5*)
				Glyma11g19570	*PDAT2*	0.9912	2.06e-06	
Glyma03g34730	Trihelix	1.83e-17	2.58e-15	*Glyma13g06080*	*BCCP2*	0.9715	6.76e-05	
				Glyma19g03530	*BCCP2*	0.9798	2.43e-05	
				Glyma09g07520	*DGAT1*	0.9940	6.63e-07	
				Glyma11g19570	*PDAT2*	0.9934	8.84e-07	
Glyma10g08370 (*ABI5*)	bZIP	4.66e-21	5.97e-09	*Glyma03g29600*	*LPAAT*	-0.9737	5.30e-05	*At2g36270* (*ABI5*)
				Glyma19g32420	*LPAAT*	-0.9831	1.43e-05	
				Glyma01g43780	*STERO*	0.9751	4.49e-05	
				Glyma08g47240	*ABI3*	0.9692	8.49e-05	
				Glyma18g38490	*ABI3*	0.9927	1.18e-06	
Glyma18g38490 (*ABI3*)	B3	9.51e-15	1.26e-01	*Glyma03g29600*	*LPAAT*	-0.9855	8.98e-06	*AT3G24650* (*ABI3*)
				Glyma19g32420	*LPAAT*	-0.9787	2.85e-05	

Co-expression network analysis was conducted between seven seed-specific TFs and genes involved in the core lipid synthesis pathways at the 1e-04 level. Group H-L1: HanDou 5 (high-oil) and ZYD4364 (low-oil); Group H-L2: HanDou 5 and Y117249 (low-oil).

doi:10.1371/journal.pone.0154882.t002

Interestingly, the expression of *WRI1* could increase seed oil content in both dicots and grasses, such as in *Arabidopsis* [20], oilseed rape [57], oil palm [58, 59] and maize [60]. However, distinct differences exist between dicots and grasses. These differences include low sequence similarity that results in dicot-specific and grass-specific clusters (S1 Dataset), different gene structures, and a low evolutionary rate in dicots (ω1 = 0.0861) as compared with that in grasses (ω0 = 0.2809) (S3 Fig). More importantly, *WRI1* is included in different regulatory networks in dicots and grasses [61]. In *Arabidopsis*, the expression of *WRI1* is up-regulated by *LEC1*, *LEC2* and *FUS3*, and *WRI1* is a direct target of *LEC2* [23], and possibly of *FUS3* [62]. In maize, however, no ortholog of *AtLEC2* was identified [60], and *WRI1* is able to regulate amino

Table 3. Soybean genes co-expressed with *GmWRI1*.

Gene			Co-expression analysis with *Glyma08g24420*		Co-expression analysis with *Glyma15g34770*		Similar result in Arabidopsis	
ID	Abbreviation	No. of AW-boxes in promoter	r	P-value	r	P-value	Homologous gene	Reference
Glyma13g06080	*BCCP2*	4	0.8105	1.53e-02	0.8491	8.24e-03	*At5g15530*	Baud et al. [20]; Maeo et al. [53]; Fukuda et al. [110]
Glyma18g50020	*BCCP2*	4	0.8844	3.93e-03	0.8771	4.66e-03	*At5g15530*	
Glyma19g03530	*BCCP2*	3	0.7445	3.36e-02	0.7842	2.17e-02	*At5g15530*	
Glyma05g36450	*BC*	1	0.8222	1.29e-02	0.7981	1.82e-02	*At5g35360*	Fukuda et al. [110]
Glyma08g03120	*BC*	1	0.9482	3.89e-04	0.9513	3.26e-04	*At5g35360*	
Glyma07g37050	*LPD*	3	0.8636	6.24e-03	0.8602	6.67e-03		
Glyma15g15310	*LPD*	1	0.8357	1.04e-02	0.8177	1.38e-02		
Glyma17g03560	*LPD*	3	0.8222	1.29e-02	0.8202	1.33e-02		
Glyma08g45990	*ENR*	2	0.8129	1.48e-02	0.8077	1.59e-02	*At2g05990*	Baud et al. [23]
Glyma18g31780	*ENR*	2	0.7291	3.90e-02	0.7354	3.68e-02	*At2g05990*	
Glyma18g36130	*FATA*	2	0.7311	3.83e-02	0.7522	3.11e-02		
Glyma07g05550	*α-PDH*	2	0.9671	1.03e-04	0.9646	1.28e-04	*At1g01090*	Baud et al. [23]
Glyma16g02090	*α-PDH*	2	0.9464	4.31e-04	0.9539	2.77e-04	*At1g01090*	
Glyma09g07520	*DGAT1*	3	0.7525	3.10e-02	0.7773	2.35e-02		
Glyma06g11860	*LACS9*	2	0.9801	2.32e-05	0.9875	5.81e-06		
Glyma18g42280	*α-CT*	1	0.7890	2.04e-02	0.7591	2.89e-02		
Glyma18g42300	*α-CT*	2	0.7972	1.84e-02	0.7664	2.67e-02		
Glyma11g19570	*PDAT2*	0	0.8025	1.71e-02	0.8144	1.45e-02		
Glyma18g06500	*MCMT*	2	0.8225	1.29e-02	0.8069	1.61e-02		

Two soybean genes (*Glyma08g24420* and *Glyma15g34770*) are homologous to *AtWRI1*, and all the genes co-expressed with the two soybean genes at the 0.05 level were listed in this table.

doi:10.1371/journal.pone.0154882.t003

acid biosynthesis [63]. In addition, *LEC1* and *FUS3* in *Arabidopsis* have low sequence similarity compared with those in grasses.

Differential expression analysis between high- and low-oil soybean accessions

A differential expression analysis to detect up-regulated genes in the high-oil materials may contribute to a better understanding of high-oil content mechanisms. In this study, RNA-seq differential expression analysis between high-oil cultivar Handou 5 (HD5; seed oil content: 22.3%) and two low oil wild soybeans ZYD4364 (11.9%) and Y117249 (12.5%) was conducted in the seeds at four developmental stages (15, 25, 35 and 55 days after flowering (DAF)). As a result, 8,356 DEGs between accessions HD5 and ZYD4364 (Group H-L1), and 5,551 DEGs between accessions HD5 and Y117249 (Group H-L2) were identified, and 3,997 common DEGs were observed ($P<0.01$).

Among 1,123 acyl-lipid metabolism genes, 77 were differentially expressed in the above two groups (H-L1 and H-L2), and 28 DEGs encoded core lipid synthetic enzymes (S2 Table and Fig 4B). We found that 28 DEGs in core lipid synthesis pathways were up-regulated in all the three accessions at stages 25 and 35 DAF, during which seed oil is rapidly accumulated (Fig 4B), indicating that they play vital roles in seed oil accumulation. Interestingly, at early (15

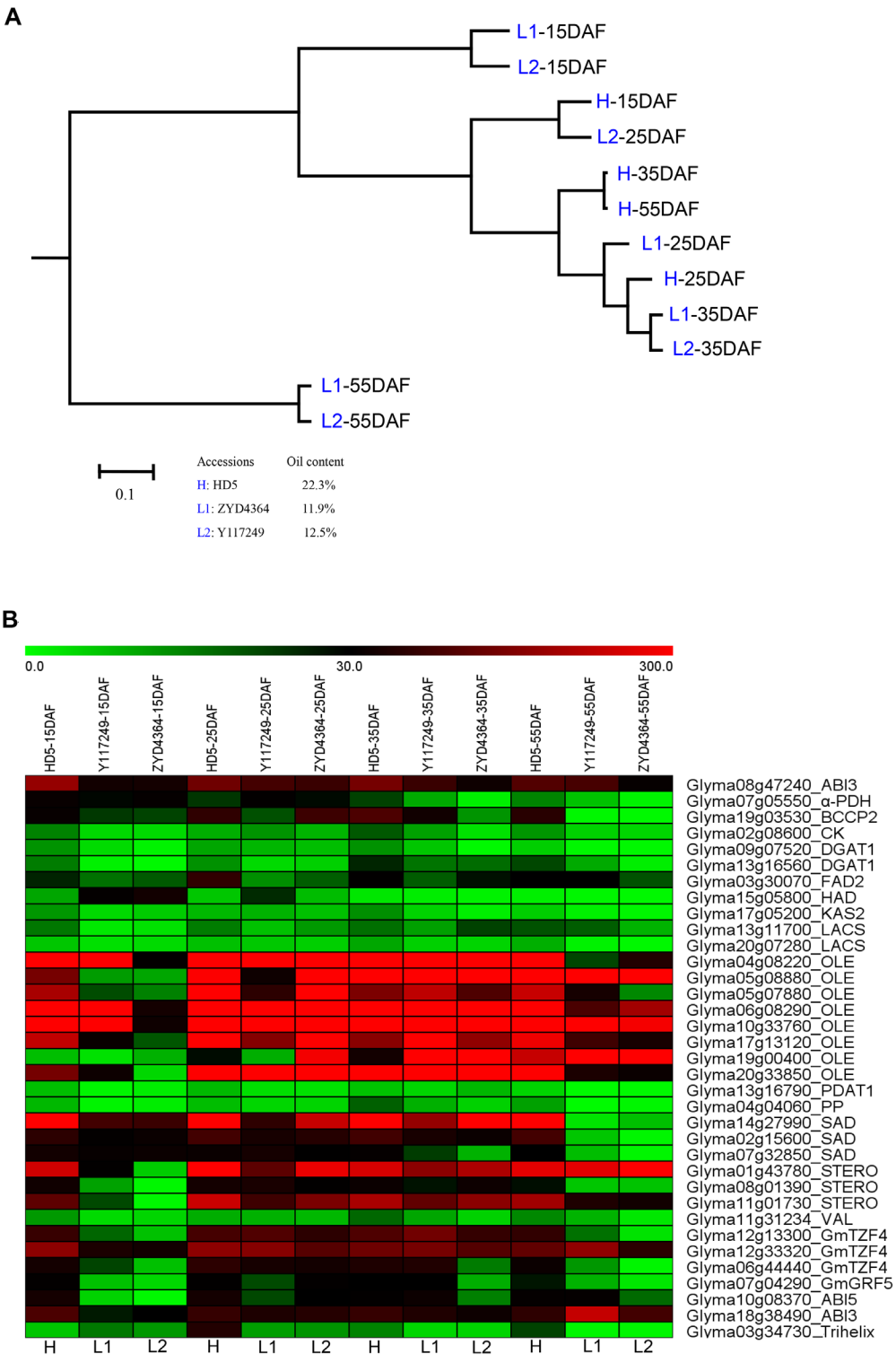

Fig 4. Expression profiles of acyl-lipid metabolism genes in high- and low-oil soybean seeds. A: Hierarchical clusters of the soybean seed samples using expression levels of 1,123 genes in acyl-lipid metabolism. 15DAF (15 days after flowering), 25DAF, 35DAF and 55DAF are four stages of seed development. B: Expression patterns of 28 differentially expressed genes in core lipid synthesis pathways. Seven seed-specific TFs are also shown. HD5 (high-oil content in seed, 22.3%), ZYD4364 (low-oil content in seed, 11.9%) and Y117249 (low-oil content in seed, 12.5%) indicate the sample codes.

doi:10.1371/journal.pone.0154882.g004

DAF) and late (55 DAF) stages, the 28 DEGs still expressed at a relative high level in high-oil accession but at a relative low level in both low-oil accessions (Fig 4B). This phenomenon was also observed in seven seed-specific lipid-related TFs. Cluster analysis of expression patterns for the 1,123 acyl-lipid metabolism genes clearly distinguished the 15 and 55 DAF from the other two stages, and the high-oil accession from the low oil accessions (Fig 4A), indicating that divergent expression patterns of acyl-lipid metabolism genes in the early and late seed development stages played roles in determining different oil content [64]. The above 28 DEGs in core lipid synthesis pathways were compared with the above 40 highly up-regulated genes. As a result, 17 common genes were found; these are shown in red in Fig 2, and encode BCCP2, HAD, SAD, PP, DGAT1, PDAT1, STERO, OLE, and ABI3. These common genes were not only up-regulated during the rapid oil accumulation phase of seed development but also differentially expressed in high- and low-oil accessions, indicating the importance of these key enzymes in determining seed oil content. Clearly, there must be a relationship between high expression pattern and key proteins. However, there are some minor exceptions; for example, *PDAT1* and *PDAT2* in *Arabidopsis* have no effects on seed oil content and TAG synthesis, respectively, although the two genes are highly expressed during seed development [65, 66].

Differential expression analysis was also used to validate the candidate lipid-related genes predicted by PLC analysis. As a result, 17 of the 93 dicot-specific lipid-related genes (Table 1 and S4 Table), and 61 of the 327 lipid-related TFs were found to be differentially expressed in the above two groups (H-L1 and H-L2) (S5 Table). More importantly, all the seven seed-specific TFs were also differentially expressed in high- and low-oil soybean accessions, except *Glyma18g38490* coding ABI3, which was differentially expressed only between accessions HD5 and ZYD4364 (Table 2).

Discussion

Carbohydrate degradation and transport

The *de novo* synthesis of FAs in plants has a close connection with carbohydrate metabolism and transportation of its metabolic intermediates [6]. Interestingly, of the 93 dicot-specific genes putatively related to lipid metabolism, 42 (45.16%) and 27 (29.03%) genes were included in carbohydrate metabolic process and transport, respectively (Table 1 and S4 Table).

The 42 genes associated with carbohydrate metabolism were classified into three sub-groups. The first one had three genes (*Glyma0165s00200*, *Glyma01g28520*, *Glyma03g08860*), encoding proteins with a conserved domain PF00686 (starch binding domain; SBD) (Table 1 and S4 Table), which is also known as the family 20 carbohydrate-binding module (CBM20) and is found in great many starch degrading enzymes including alpha-amylase, beta-amylase, glucoamylase, and cyclodextrin glucanotransferase [67]. SBD could mediate the attachment between starch-active enzymes and starch granules and might disrupt the structure of the starch surface, thereby enhancing the amylolytic rate [68–70]. Accordingly, these three CBM20-like proteins possibly promote the hydrolysis of starch by combining with some amylolytic enzymes or other related glycosidases. Although all the three CBM20-like genes were not differentially expressed at the 0.01 probability level, *Glyma01g28520* was differentially expressed at the 0.05 level (S4 Table).

The second sub-group includes the dicot-specific genes that participate in the degradation of other carbohydrates, such as homogalacturonan degradation (PWY-1081) and chitin degradation II (PWY-6902) (Table 1 and S4 Table). Note that *Glyma09g04330* in chitin degradation II was differentially expressed between high- and low-oil soybean accessions. These carbohydrates degradation genes might be involved in the increase in carbon sources during FA biosynthesis. The last sub-group comprised some dicot-specific genes with unknown functions (Table 1 and S4 Table).

Among the 27 dicot-specific genes in transport, four genes (*Glyma07g38830*, *Glyma13g27680*, *Glyma15g11270* and *Glyma18g08740*) belonged to *PF03151* (Triose-phosphate Transporter family). The first three genes were annotated as genes encoding the glucose 6-phosphate/phosphate translocator 2 (GPT2), which is located in the envelope of the plastids for the importation of Glc6P [71]. In non-green plastids, glucose 6-phosphate (Glc6P) can be used as a precursor for starch biosynthesis and FA synthesis and also as a substrate for the Oxidative Pentose Phosphate Pathway (OPPP) which supplies reducing power to drive FA synthesis [4, 72]. Although *AtGPT2* has no obvious effects on plant development [73, 74], its over-expression could increase the net import of Glc6P from cytosol to chloroplast and accelerate the accumulation of soluble sugars in *Arabidopsis* [75, 76]. Moreover, the expression of *GPT2* in oil palm increased notably during fruit ripening and was significantly higher than in date palm, a low oil content palm [77]. More importantly, *Glyma15g11270* coding GPT2 was also differentially expressed in high- and low-oil soybean accessions (Table 1 and S4 Table). In that case, the specific GPT2 in high-oil content dicots implies a strong funneling of carbon toward pyruvate in the plastid, a significant increase for FA synthesis and ultimately an increase in the seed oil content [77]. The remaining 23 transport genes had unclear functions (Table 1 and S4 Table).

FA synthesis and regulation of seed oil accumulation

Functions of dicot-specific genes encoding key enzymes of FA synthesis. Among the 23 dicot-specific genes encoding key enzymes of FA synthesis in clusters 3 to 5 (S2 Table), 10 are involved in FA synthesis, i.e., *Glyma05g36450* and *Glyma08g03120* encode Biotin Carboxylase (*BC*); *Glyma13g06080*, *Glyma18g50020* and *Glyma19g03530* encode Biotin Carboxyl Carrier Protein 2 (*BCCP2*); *Glyma18g42280* and *Glyma18g42300* encode Carboxyltransferase alpha subunit (*α-CT*); *Glyma15g34770* and *Glyma08g24420* belong to the *AP2/EREBP* family and encode soybean *WRI1*; and *Glyma16g05480* encodes the FUSCA3 (FUS3). All these genes except the two genes encoding BC were highly up-regulated during the rapid oil accumulation phase of seed development. *Glyma19g03530* encoding BCCP2 was up-regulated in high-oil soybean accessions compared with the two low-oil accessions (Fig 2).

BC, *BCCP* and *α-CT* are the subunits of heteromeric acetyl-CoA carboxylase (ACCase), which is the rate limiting enzyme of FA synthesis [78]. In dicots, there are two forms of Acetyl-CoA Carboxylase (ACCase), a heteromeric form in the plastid and a homomeric form in the cytosol. In grasses, however, ACCase in both the plastid and the cytosol is the homomeric type [79]. Previous studies have shown that ACCase is able to control the rate of carbon flow in plant leaves and its expression level is related to oil content in seed [9,11]. Specific heteromeric ACCase and its high expression level in high-oil dicots are probably key factors leading to the fact that oil content in high-oil dicots is significantly higher than that in grasses.

WRI1 plays an important role in controlling the rate of carbon flow from carbohydrate metabolism to lipid synthesis, and is capable of affecting the seed oil accumulation by regulating a set of genes involved in lipid synthesis, glycolysis and photosynthesis [80, 57, 47]. Although *WRI1* could regulate seed oil accumulation in both grasses and high-oil dicots, different gene structures and divergent evolutionary rates between these two lineages were observed (S3 Fig). We deduced that different evolutionary mechanisms of *WRI1* in high-oil dicots and grasses might lead to different regulatory networks [61].

Although up-regulation of key enzymes in lipid synthesis like ACCase alone slightly increases oil content [11], its effect on increasing the oil content was much less than that by up-regulation of *WRI1* or other TFs [47, 59]. This partly indicates that oil accumulation is a complex biological process and increasing the expression of TFs may be an effective approach to significantly improve seed oil content.

Functions for up-regulated genes of FA synthesis in seed. In the initial period of FA synthesis, apart from five dicot-specific genes coding ACCase, five genes coding PDHC were also found to be highly up-regulated during the rapid oil accumulation phase of seed development (Fig 2). In several species, the expression of *PDHC* has been reported to be associated with seed oil content [81–83]. In this study, *Glyma07g05550* encoding a subunit of PDHC was found to be a DEG between high- and low-oil soybean accessions (Fig 2). A similar phenomenon between high- and low-oil accessions in oat was also observed by Hayden et al. [84]. Therefore, we assumed that the up-regulated expression of *PDHC* and *ACCase* in soybean likely resulted in the increase of carbon flux to FA synthesis, and then to an increase of the efficiency of FA synthesis [84].

FA transportation, triacylglycerol synthesis and oil-body formation

FA transportation. Among the 54 dicot-specific genes involved in acyl-lipid metabolism, 12 genes encoded lipid transfer proteins (LTPs), including 5 DEGs between high- and low-oil soybean accessions (S2 Table). Apart from the above 12 high-oil dicot-specific LTPs, we also identified nine additional DEGs coding LTPs (S2 Table). Wang et al. [64] hypothesizes that an increase of the number of *LTP1* genes in sesame might enhance oil accumulation by strengthening the transport of FAs, acyl-CoAs, and other lipid molecules. On this basis, we proposed that abundant *LTP* genes in oil plants might possibly benefit oil accumulation. Among the 40 up-regulated genes during the stages from 5–6 mg to 400–500 mg (Fig 2), two genes *Glyma08g01180* and *Glyma06g11860* encoded ACP4 and LACS9, respectively. In addition, two genes (*Glyma13g11700* and *Glyma20g07280*) that belong to LACS family are up-regulated DEGs in high-oil accession (Fig 4B). Therefore, dicot-specific or up-regulation of *LTP*, *ACP4* and *LACS* in dicots could possibly increase the efficiency of plastidial fatty acid export for TAG synthesis and then consequently regulate seed oil content [55, 85, 86].

Triacylglycerol synthesis and oil-body formation. In TAG assembly, eight genes were highly up-regulated. Among these eight, *Glyma04g04060* encoding PP, *Glyma09g07520* and *Glyma13g16560 en*coding DGAT1, and *Glyma13g16790* encoding PDAT1 were also differentially expressed in high- and low-oil soybean accessions (Fig 2). *DGAT1* has a principal role in TAG biosynthesis [87] and over-expression of *DGAT1* had been shown to enhance oil accumulation [16, 32, 88]. In *B. napus*, two domestication-related genes *BnaA01g32210D* and *BnaAnng30990D*, encoding PP and DGAT1, respectively [89], possibly played roles in increasinging seed oil content. In oil-body formation, three and six genes respectively encoding STEROs and OLEs were identified and were up-regulated by some hundredfold between the stages from 5–6 mg to 400–500 mg (Fig 3 and S4 Dataset). Phylogenetic analysis showed more copies of genes coding OLEs in the dicots than in the grasses (Fig 3). Highly up-regulated and abundant OLEs possibly play vital roles in regulating seed oil content in dicots. Evidently, Siloto et al. [90] and Miquel et al. [51] showed that the size and spatial distribution of oil bodies affect the total lipid content and oil body proteins have specific functions in lipid accumulation.

We also found that 40 highly up-regulated genes and 28 DEGs in high- and low-oil soybean accessions were all enriched significantly in downstream section of the TAG biosynthesis pathway and the oil-body formation pathway (Figs 2 and 4B). This phenomenon was also observed in sesame [64]. Therefore, we deduced that enzymes in the downstream section of the TAG biosynthesis pathway and the oil-body formation pathway played vital roles in the variation of seed oil content in soybean and sesame [64]. In other words, the efficient flow of fatty acids to formation of TAG and oil-body may ultimately influence seed oil content.

Signal transduction and other factors involved in lipid metabolism

Among the 54 dicot-specific genes involved in acyl-lipid metabolism (S2 Table), *Glyma07g01310*, *Glyma08g20710* and *Glyma15g02710* encode Phospholipase Dα (*PLDα*), which is able to hydrolyze phospholipids, producing signalling molecule phosphatidic acid [91]. A suppression of the expression of *PLDα* led to a significant decrease in triacylglycerol levels in *Arabidopsis* leaves [26] and could also slow the conversion of phosphatidylcholine to TAG in soybean seeds [92]. Two transgenic *B. napus* cultivars expressing an *Arabidopsis PLDα1* both demonstrated a 9% increase in seed total oil content [93]. In addition, *PLDα* is regulated by *GmMYB73*, which has been shown to be an important TF regulating lipid content [26]. Although these three dicot-specific *PLDα* are not DEGs in high- and low-oil soybean accessions, two of their homologous genes, *Glyma13g44170* and *Glyma08g22600*, are found to be differentially expressed genes (S2 Table). We suspect that up-regulated dicot-specific *PLDα* and the other two DEGs of *PLDα* would possibly have a significant effect on the seed oil content.

Among the 93 dicot-specific genes associated with acyl-lipid metabolism, 11 genes were in the category GO:0007165 (signal transduction) (Table 1 and S4 Table). Among these 11 genes, three (*Glyma09g07090*, *Glyma15g18380* and *Glyma17g06290*) encode GATA TFs, which play a role in light-mediated transcriptional regulation [94]. The three genes were homologous to the *Arabidopsis* gene *At5g56860* that is closely associated with lipid metabolism in green algae [95] and is capable of regulating carbon and nitrogen metabolism [96]. In this study, *Glyma15g18380* was validated to be a DEG in high- and low-oil soybean accessions. We deduced that these specific GATA TFs might affect the accumulation of seed oil in dicots. Two genes, *Glyma09g39570* and *Glyma10g12130*, are included in category PWY-5035 (gibberellin biosynthesis III) and might be related to the synthesis of the hormone gibberellin. Similarly, five dicot-specific genes, *Glyma04g02660*, *Glyma06g02690*, *Glyma14g40400*, *Glyma17g37750* and *Glyma17g37760* in GO:0005975 (carbohydrate metabolism) were annotated as PF02704 (Gibberellin regulated protein). The expression of genes encoding gibberellin regulated proteins is up-regulated by gibberellin [97]. It is known that the gibberellin signalling pathway is related to FA content [98]. More importantly, four genes (*Glyma04g02660*, *Glyma06g02690*, *Glyma17g37750* and *Glyma17g37760*) encoding gibberellin regulated proteins were also validated to be DEGs in high- and low-oil soybean accessions. Therefore, the above six gibberellin related genes may play roles in FA synthesis and thereby affect the oil content.

Other factors involved in lipid metabolism. Apart from the above specific genes directly participating in acyl-lipid metabolism, we also found that five dicot-specific genes in soybean were putatively involved in the lipid metabolic process (GO:0006629) (Table 1 and S4 Table). Among these genes, *Glyma02g45680* and *Glyma14g03130* encoded cytochrome P450, which plays a role in the FA metabolism in plants [99, 100]); and *Glyma18g45250* encoded a protein participating in glyceollin biosynthesis, which was related to lipid peroxidation [101]. Since these genes were highly expressed and coordinated with genes participating in acyl lipid synthesis, we hypothesized that these genes and their associated pathways possibly play some roles in seed oil accumulation.

Conclusion

Ninety-three dicot-specific genes, including 42 and 27 genes respectively in carbohydrate degradation and transport, were predicted to be candidate genes associated with acyl-lipid metabolism pathways. And seed-specific TF genes *GmGRF5*, *ABI5* and *GmTZF4* were also predicted to play roles in regulating seed oil accumulation. Furthermore, *ACCase*, *DGAT1*, *PP*, *OLEs* and *STEROs* were highly up-regulated not only in specific stages of seed development but also in

high-oil accessions, which indicates that enzymes in initial fatty acid synthesis, downstream of TAG biosynthesis and oil-body formation, played vital roles in the variation of seed oil content. In particular, highly up-regulated and abundant *OLEs* possibly play vital roles in determining seed oil content. Most of the above key genes were further confirmed by differential gene expression analysis between high-oil cultivated and low-oil wild soybeans.

Materials and Methods

Genomic data and gene expression data

The genomic data of four high-oil dicot species (*Glycine max* (seed oil content: 20%), *Gossypium raimondii* (30%), *Ricinus communis* (50%), and *Arabidopsis thaliana* (35%)) and three low-oil grasses (*Sorghum bicolor* (3%), *Setaria italic* (1.7%), and *Oryza sativa* (3%)) were downloaded from Phytozome V9.1 (http://www.phytozome.net/) [102]. The longest encoded protein sequence was chosen for genes with multiple transcripts.

The transcriptome data with two biological replicates of *G. max* Williams 82 [41] were downloaded from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42871). The data included seven stages of seed development: whole seed 4 DAF, whole seed 12–14 DAF, whole seed 22–24 DAF, whole seed 5–6 mg in weight, cotyledons 100–200 mg in weight, cotyledon 400–500 mg in weight, and dry whole seed.

GO annotation and GO enrichment analysis

The GO annotations of the *G. max* genes, including molecular function, molecular location and biological process, were conducted using the online tool Goanna (http://agbase.msstate.edu/cgi-bin/tools/GOanna.cgi) [103]. The GO enrichment analysis was performed using GOstats with a threshold P value of less than 0.01 [104]. The GO slims, which are a subset of GO terms for a high level summary of the ontology content, were summarized using GOSlimViewer (http://www.agbase.msstate.edu/cgi-bin/tools/goslimviewer_select.pl) [103].

OrthoMCL analysis and definition of lineage-specific clusters

Orthologous gene clusters were calculated from OrthoMCL comparisons of four dicots and three Grasses [40].

Based on all-against-all BLASTP comparisons of a set of protein sequences from genomes of interest, clusters of proteins were grouped according to reciprocal best similarity pairs between and within species, using OrthoMCL software implemented by the Markov Clustering algorithm (MCL; http://micans.org/mcl/) [105]. Here all the critical values were set as default values in the software. One OG was defined as dicot-specific if all the genes in the OG were coming from all the four rosid species but not from any grass species.

Prediction of genes involved in lipid metabolism of *Glycine max*

More than 600 genes involved in acyl-lipid metabolism in Arabidopsis [7,8] were downloaded from the website ARALIP (http://aralip.plantbiology.msu.edu/). Based on the OrthoMCL result, all the orthologous and paralogous genes participating in acyl lipid biosynthesis in soybean were identified.

Clustering by expression pattern and PLC network analysis

Using the transcriptome data, gene models were clustered using BioLayout Express^{3D}, implemented by the MCL [42]; and any two genes with an absolute value of the Pearson correlation coefficient greater than 0.7 were considered to be similar in their expression level [43, 106].

The PLC network analysis developed by Wei et al. [43] was used to identify new candidate pathway members in the lipid synthesis pathway. Genes co-expressed with more than two genes in the soybean acyl-lipid metabolism pathway were considered as candidate lipid synthesis pathway members. A relatively stringent correlation threshold was set at the probability value of 1e-4.

RNA-seq for transcriptome analysis of high- and low-oil soybean accessions

The materials used for RNA-seq to analyze lipid synthesis were three soybean accessions, one high-oil cultivar Handou 5 (HD5; seed oil content: 22.3%) and two low oil wild soybeans ZYD4364 (11.9%) and Y117249 (12.5%). Whole seeds at stages 15, 25, 35 and 55 DAF were harvested as samples. Total RNA of every sample was extracted from tissues using the TRIzol reagent (Invitrogen). The quality and the concentration of total RNA were quantified separately using Agilent 2200 TapeStation and ND-1000 Nanodrop. cDNA library were constructed using the same procedure described in Severin et al. [107] and sequenced using an Illumina HiSeq 2500 sequencing platform.

The raw reads were cleaned by removing reads with adapters and those of low quality. Clean reads were mapped to reference sequences using SOAPaligner/soap2 (http://soap.genomics.org.cn/soapdenovo.html). Mismatches no more than two bases were allowed in the alignment. The gene expression level was calculated by using RPKM method (Reads Per kb per Million reads) [108]. Fisher's exact test method in DEGseq package [109] was used to identify DEGs between high- and low-oil accessions, at the 0.01 significant level.

Supporting Information

S1 Dataset. OGs of all the genes and genes in acyl-lipid metabolism of seven species. Dicot-specific OGs shared with all the four dicots and 4051 dicots-specific genes in soybean are also shown.
(XLSX)

S2 Dataset. GO enrichment analysis in *Glycine max* for 4051 soybean genes in dicot-specific OGs. GOslim summary of the enriched GO terms are also listed.
(XLSX)

S3 Dataset. GO enrichment analysis for 805 dicot-specific genes of clusters 3 to 5 in *Glycine max*. GOslim summary of the enriched GO terms are also listed.
(XLSX)

S4 Dataset. Up-regulated genes of lipid synthesis core pathways in at least one of the seed developmental stages from 5–6 mg to 400–500 mg.
(XLSX)

S5 Dataset. Predicted dicot-specific genes and TFs that are associated with acyl-lipid metabolism and their co-expression network with acyl-lipid metabolism genes in soybean. Co-expression network of 19 known lipid-related transcription factors with core enzymes in lipid synthesis ($P<0.05$) are also shown.
(XLSX)

S1 Fig. Co-expression network among soybean genes coding *ABI3*, *OLE* and *STERO*. Two genes with a coordinated relationship were linked by regular (P-value < 0.01) or bold (P-value $<$ 1e-04) lines.
(PDF)

S2 Fig. AW-boxes, with arrowhead, in 5'-upstream sequences of *GmWRI1* coordinated genes involved in lipid synthesis pathways. Schematic drawing of the sequence 2 kb upstream of the ATG start codon of *GmWRI1* co-expressed genes.
(PDF)

S3 Fig. Phylogenetic analysis of *WRI1*. A: Phylogenetic tree of *WRI1* constructed by MEGA 6.0 using Neighbor-Joining method and the bootstrap test was performed with 1,000 iterations. Square boxes indicate duplication events and numbers on the branches represent the bootstrap support. Genes structure visualizing positions of exons and introns are also shown; this was constructed by GSDS 2.0. B: Selection detection using branch model implemented by PAML.
(PDF)

S1 Table. Genomic information of the seven species used in this study.
(PDF)

S2 Table. Acyl-lipid metabolism genes that are high-oil dicot-specific or differentially expressed genes in high- and low-oil soybean accessions.
(PDF)

S3 Table. 207 genes in interested biological processes that were included in the 805 genes in clusters 3 to 5.
(PDF)

S4 Table. Annotation and differential expression analysis of 93 dicot-specific genes that were associated with oil accumulation.
(PDF)

S5 Table. 327 lipid-related TFs predicted by PLC analysis.
(PDF)

Acknowledgments

We thank Dr Guo Liang (Huazhong Agricultural University) for helpful comments on the revised version of this manuscript.

Author Contributions

Conceived and designed the experiments: YMZ. Performed the experiments: LZ SBW QGL JS YQH LZ YMZ. Analyzed the data: LZ JS YQH. Wrote the paper: LZ YMZ HQZ JMD.

References

1. Scotland RW, Wortley AH. How many species of seed plants are there? Taxon. 2003; 52(1):101.
2. Raven P, Johnson G, Mason K, Losos J, Singer S. Biology: McGraw-Hill Education, 2010.
3. Bao X, Focke M, Pollard M, Ohlrogge J. Understanding *in vivo* carbon precursor supply for fatty acid synthesis in leaf tissue. Plant J. 2000; 22(1):39–50. PMID: 10792819
4. Rawsthorne S. Carbon flux and fatty acid synthesis in plants. Prog Lipid Res. 2002; 41(2):182–96. PMID: 11755683
5. Li Y, Han D, Hu G, Sommerfeld M, Hu Q. Inhibition of starch synthesis results in overproduction of lipids in *Chlamydomonas reinhardtii*. Biotechnol Bioengineer. 2010; 107(2):258–68.
6. Weselake RJ, Taylor DC, Rahman MH, Shah S, Laroche A, McVetty PB et al. Increasing the flow of carbon into seed oil. Biotechnol Adv. 2009; 27(6):866–78. doi: 10.1016/j.biotechadv.2009.07.001 PMID: 19625012
7. Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD et al. Acyl-lipid metabolism. The Arabidopsis book / American Society of Plant Biologists. 2013; 11:e0161. doi: 10.1199/tab.0161 PMID: 23505340

8. Beisson F, Koo AJ, Ruuska S, Schwender J, Pollard M, Thelen JJ et al. Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. Plant Physiol. 2003; 132(2):681–97. PMID: 12805597
9. Klaus D, Ohlrogge JB, Neuhaus HE, Dormann P. Increased fatty acid production in potato by engineering of acetyl-CoA carboxylase. Planta. 2004; 219(3):389–96. PMID: 15014998
10. Roesler K, Shintani D, Savage L, Boddupalli S, Ohlrogge J. Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiol. 1997; 113(1):75–81. PMID: 9008389
11. Davis MS, Solbiati J, Cronan JE Jr. Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. J Biol Chem. 2000; 275(37):28593–8. PMID: 10893421
12. Vigeolas H, Waldeck P, Zank T, Geigenberger P. Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. Plant Biotechnol J. 2007; 5(3):431–41. PMID: 17430545
13. Jain RK, Coffey M, Lai K, Kumar A, MacKenzie SL. Enhancement of seed oil content by expression of glycerol-3-phosphate acyltransferase genes. Biochem Soc Trans. 2000; 28:958–61. PMID: 11171271
14. Maisonneuve S, Bessoule JJ, Lessire R, Delseny M, Roscoe TJ. Expression of rapeseed microsomal lysophosphatidic acid acyltransferase isozymes enhances seed oil content in Arabidopsis. Plant Physiol. 2010; 152(2):670–84. doi: 10.1104/pp.109.148247 PMID: 19965969
15. Jako C, Kumar A, Wei Y, Zou J, Barton DL, Giblin EM et al. Seed-specific over-expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiol. 2001; 126(2):861–74. PMID: 11402213
16. Kim H, Kim HU, Suh MC. Efficiency for increasing seed oil content using *WRINKLED1* and *DGAT1* under the control of two seed-specific promoters, FAE1 and Napin. J Plant Biotechnol. 2012; 39 (4):242–52.
17. Lardizabal K, Effertz R, Levering C, Mai J, Pedroso MC, Jury T et al. Expression of *Umbelopsis ramanniana DGAT2A* in seed increases oil in soybean. Plant Physiol. 2008; 148(1):89–96. doi: 10.1104/pp.108.123042 PMID: 18633120
18. Weselake RJ, Shah S, Tang M, Quant PA, Snyder CL, Furukawa-Stoffer TL et al. Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. J Exp Bot. 2008; 59(13):3543–9. doi: 10.1093/jxb/ern206 PMID: 18703491
19. Zheng P, Allen WB, Roesler K, Williams ME, Zhang S, Li J et al. A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nat Genet. 2008; 40(3):367–72. doi: 10.1038/ng.85 PMID: 18278045
20. Baud S, Wuilleme S, To A, Rochat C, Lepiniec L. Role of WRINKLED1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in Arabidopsis. Plant J. 2009; 60(6):933–47. doi: 10.1111/j.1365-313X.2009.04011.x PMID: 19719479
21. Mu JY, Tan HL, Zheng Q, Fu FY, Liang Y, Zhang JA et al. *LEAFY COTYLEDON1* is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant Physiol. 2008; 148(2):1042–54. doi: 10.1104/pp.108.126342 PMID: 18689444
22. Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J et al. Enhanced seed oil production in canola by conditional expression of *Brassica napus LEAFY COTYLEDON1* and *LEC1-LIKE* in developing seeds. Plant Physiol. 2011; 156(3):1577–88. doi: 10.1104/pp.111.175000 PMID: 21562329
23. Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. Plant J. 2007; 50(5):825–38. PMID: 17419836
24. Wang H, Guo J, Lambert KN, Lin Y. Developmental control of *Arabidopsis* seed oil biosynthesis. Planta. 2007; 226(3):773–83. PMID: 17522888
25. Song QX, Li QT, Liu YF, Zhang FX, Ma B, Zhang WK et al. Soybean *GmbZIP123* gene enhances lipid content in the seeds of transgenic *Arabidopsis* plants. J Exp Bot. 2013; 64(14): 4329–41. doi: 10.1093/jxb/ert238 PMID: 23963672
26. Liu YF, Li QT, Lu X, Song QX, Lam SM, Zhang WK et al. Soybean *GmMYB73* promotes lipid accumulation in transgenic plants. BMC Plant Biol. 2014; 14:73. doi: 10.1186/1471-2229-14-73 PMID: 24655684
27. Crowe AJ, Abenes M, Plant A, Moloney MM. The seed-specific transactivator, ABI3, induces oleosin gene expression. Plant Sci. 2000; 151(2):171–81. PMID: 10808073

28. Monke G, Seifert M, Keilwagen J, Mohr M, Grosse I, Hahnel U et al. Toward the identification and regulation of the *Arabidopsis thaliana* ABI3 regulon. Nucleic Acids Res. 2012; 40(17):8240–54. PMID: 22730287

29. Smooker AM, Wells R, Morgan C, Beaudoin F, Cho K, Fraser F et al. The identification and mapping of candidate genes and QTL involved in the fatty acid desaturation pathway in *Brassica napus*. Theor Appl Genet. 2011; 122(6):1075–90. doi: 10.1007/s00122-010-1512-5 PMID: 21184048

30. Ying JZ, Shan JX, Gao JP, Zhu MZ, Shi M, Lin HX. Identification of quantitative trait loci for lipid metabolism in rice seeds. Mol Plant. 2012; 5(4):865–75. doi: 10.1093/mp/ssr100 PMID: 22147755

31. Schwender J, Hay JO. Predictive modeling of biomass component tradeoffs in *Brassica napus* developing oilseeds based on in silico manipulation of storage metabolism. Plant Physiol. 2012; 160 (3):1218–36. doi: 10.1104/pp.112.203927 PMID: 22984123

32. van Erp H, Kelly AA, Menard G, Eastmond PJ. Multigene engineering of triacylglycerol metabolism boosts seed oil content in Arabidopsis. Plant Physiol. 2014; 165(1):30–6. doi: 10.1104/pp.114.236430 PMID: 24696520

33. Andre C, Froehlich JE, Moll MR, Benning C. A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in *Arabidopsis*. Plant Cell. 2007; 19(6):2006–22. PMID: 17557808

34. Andriotis VM, Kruger NJ, Pike MJ, Smith AM. Plastidial glycolysis in developing Arabidopsis embryos. The New Phytologist. 2010; 185(3):649–62. doi: 10.1111/j.1469-8137.2009.03113.x PMID: 20002588

35. Allen DK, Young JD. Carbon and nitrogen provisions alter the metabolic flux in developing soybean embryos. Plant Physiol. 2013; 161(3):1458–75. doi: 10.1104/pp.112.203299 PMID: 23314943

36. Wakao S, Andre C, Benning C. Functional analyses of cytosolic glucose-6-phosphate dehydrogenases and their contribution to seed oil accumulation in *Arabidopsis*. Plant Physiol. 2008; 146(1):277–88. PMID: 17993547

37. Sanjaya, Durrett TP, Weise SE, Benning C. Increasing the energy density of vegetative tissues by diverting carbon from starch to oil biosynthesis in transgenic Arabidopsis. Plant Biotechnol J. 2011; 9 (8):874–83. PMID: 22003502

38. Meyer K, Stecca KL, Ewell-Hicks K, Allen SM, Everard JD. Oil and protein accumulation in developing seeds is influenced by the expression of a cytosolic pyrophosphatase in *Arabidopsis*. Plant Physiol. 2012; 159(3):1221–34. doi: 10.1104/pp.112.198309 PMID: 22566496

39. Pedersen P, Elbert B. Soybean growth and development. Iowa State University, University Extension Ames, IA; 2004.

40. Li L, Stoeckert CJ Jr., Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003; 13(9):2178–89. PMID: 12952885

41. Jones SI, Vodkin LO. Using RNA-Seq to profile soybean seed development from fertilization to maturity. PLoS ONE. 2013; 8(3):e59270. doi: 10.1371/journal.pone.0059270 PMID: 23555009

42. Theocharidis A, van Dongen S, Enright AJ, Freeman TC. Network visualization and analysis of gene expression data using BioLayout Express(3D). Nature Protocols. 2009; 4(10):1535–50. doi: 10.1038/nprot.2009.177 PMID: 19798086

43. Wei H, Persson S, Mehta T, Srinivasasainagendra V, Chen L, Page GP et al. Transcriptional coordination of the metabolic network in *Arabidopsis*. Plant Physiol. 2006; 142(2):762–74. PMID: 16920875

44. Mochida K, Uehara-Yamaguchi Y, Yoshida T, Sakurai T, Shinozaki K. Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticeae crops. Plant Cell Physiol. 2011; 52(5):785–803. doi: 10.1093/pcp/pcr035 PMID: 21441235

45. Guo K, Zou W, Feng Y, Zhang M, Zhang J, Tu F et al. An integrated genomic and metabolomic framework for cell wall biology in rice. BMC Genomics. 2014; 15:596. doi: 10.1186/1471-2164-15-596 PMID: 25023612

46. Fu FF, Xue HW. Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. Plant Physiol. 2010; 154(2):927–38. doi: 10.1104/pp.110.159517 PMID: 20713616

47. Baud S, Lepiniec L. Regulation of de novo fatty acid synthesis in maturing oilseeds of *Arabidopsis*. Plant Physiol Biochem. 2009; 47(6):448–55. doi: 10.1016/j.plaphy.2008.12.006 PMID: 19136270

48. Banas W, Sanchez Garcia A, Banas A, Stymne S. Activities of acyl-CoA:diacylglycerol acyltransferase (DGAT) and phospholipid:diacylglycerol acyltransferase (PDAT) in microsomal preparations of developing sunflower and safflower seeds. Planta. 2013; 237(6):1627–36. doi: 10.1007/s00425-013-1870-8 PMID: 23539042

49. Kim HU, Hsieh K, Ratnayake C, Huang AH. A novel group of oleosins is present inside the pollen of Arabidopsis. J Biol Chem. 2002; 277(25):22677–84. PMID: 11929861

50. Liu WX, Liu HL, Qu le Q. Embryo-specific expression of soybean oleosin altered oil body morphogenesis and increased lipid content in transgenic rice seeds. Theor Appl Genet. 2013; 126(9):2289–97. doi: 10.1007/s00122-013-2135-4 PMID: 23748707

51. Miquel M, Trigui G, d'Andrea S, Kelemen Z, Baud S, Berger A et al. Specialization of oleosins in oil body dynamics during seed development in Arabidopsis seeds. Plant Physiol. 2014; 164(4):1866–78. doi: 10.1104/pp.113.233262 PMID: 24515832

52. Wang HW, Zhang B, Hao YJ, Huang J, Tian AG, Liao Y et al. The soybean Dof-type transcription factor genes, *GmDof4* and *GmDof11*, enhance lipid content in the seeds of transgenic Arabidopsis plants. Plant J. 2007; 52(4):716–29. PMID: 17877700

53. Maeo K, Tokuda T, Ayame A, Mitsui N, Kawai T, Tsukagoshi H et al. An AP2-type transcription factor, WRINKLED1, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. Plant J. 2009; 60(3):476–87. doi: 10.1111/j.1365-313X.2009.03967.x PMID: 19594710

54. Jin J, Zhang H, Kong L, Gao G, Luo J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. 2014; 42:D1182–7. doi: 10.1093/nar/gkt1016 PMID: 24174544

55. Jessen D, Roth C, Wiermer M, Fulda M. Two activities of long-chain acyl-coenzyme A synthetase are involved in lipid trafficking between the endoplasmic reticulum and the plastid in Arabidopsis. Plant Physiol. 2015; 167(2):351–66. doi: 10.1104/pp.114.250365 PMID: 25540329

56. Zhao L, Katavic V, Li F, Haughn GW, Kunst L. Insertional mutant analysis reveals that long-chain acyl-CoA synthetase 1 (LACS1), but not LACS8, functionally overlaps with LACS9 in Arabidopsis seed oil biosynthesis. Plant J. 2010; 64(6):1048–58. doi: 10.1111/j.1365-313X.2010.04396.x PMID: 21143684

57. Wu XL, Liu ZH, Hu ZH, Huang RZ. *BnWRI1* coordinates fatty acid biosynthesis and photosynthesis pathways during oil accumulation in rapeseed. J Integr Plant Biol. 2014; 56(6):582–93. doi: 10.1111/jipb.12158 PMID: 24393360

58. Ma W, Kong Q, Arondel V, Kilaru A, Bates PD, Thrower NA et al. *Wrinkled1*, a ubiquitous regulator in oil accumulating tissues from *Arabidopsis* embryos to oil palm mesocarp. PLoS ONE. 2013; 8(7): e68887. doi: 10.1371/journal.pone.0068887 PMID: 23922666

59. Dussert S, Guerin C, Andersson M, Joet T, Tranbarger TJ, Pizot M et al. Comparative transcriptome analysis of three oil Palm fruit and seed tissues that differ in oil content and fatty acid composition. Plant Physiol. 2013; 162(3):1337–58. doi: 10.1104/pp.113.220525 PMID: 23735505

60. Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J et al. Expression of *ZmLEC1* and *ZmWRI1* increases seed oil production in maize. Plant Physiol. 2010; 153(3):980–7. doi: 10.1104/pp.110.157537 PMID: 20488892

61. Sreenivasulu N, Wobus U. Seed-development programs: a systems biology-based comparison between dicots and monocots. Ann Rev Plant Biol. 2013; 64:189–217.

62. Yamamoto A, Kagaya Y, Usui H, Hobo T, Takeda S, Hattori T. Diverse roles and mechanisms of gene regulation by the Arabidopsis seed maturation master regulator FUS3 revealed by microarray analysis. Plant Cell Physiol. 2010; 51(12):2031–46. doi: 10.1093/pcp/pcq162 PMID: 21045071

63. Pouvreau B, Baud S, Vernoud V, Morin V, Py C, Gendrot G et al. Duplicate maize *Wrinkled1* transcription factors activate target genes involved in seed oil biosynthesis. Plant Physiol. 2011; 156(2):674–86. doi: 10.1104/pp.111.173641 PMID: 21474435

64. Wang L, Yu S, Tong C, Zhao Y, Liu Y, Song C et al. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol. 2014; 15(2):R39. doi: 10.1186/gb-2014-15-2-r39 PMID: 24576357

65. Mhaske V, Beldjilali K, Ohlrogge J, Pollard M. Isolation and characterization of an *Arabidopsis thaliana* knockout line for phospholipid: diacylglycerol transacylase gene (*At5g13640*). Plant Physiol Biochem. 2005; 43(4):413–7. PMID: 15907694

66. van Erp H, Bates PD, Burgal J, Shockey J, Browse J. Castor phospholipid:diacylglycerol acyltransferase facilitates efficient metabolism of hydroxy fatty acids in transgenic Arabidopsis. Plant Physiol. 2011; 155(2):683–93. doi: 10.1104/pp.110.167239 PMID: 21173026

67. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015; 43:D222–6. doi: 10.1093/nar/gku1221 PMID: 25414356

68. Christiansen C, Abou Hachem M, Janecek S, Vikso-Nielsen A, Blennow A, Svensson B. The carbohydrate-binding module family 20—diversity, structure, and function. The FEBS Journal. 2009; 276 (18):5006–29. doi: 10.1111/j.1742-4658.2009.07221.x PMID: 19682075

69. Rodriguez-Sanoja R, Oviedo N, Sanchez S. Microbial starch-binding domain. Curr Opinion Microbiol. 2005; 8(3):260–7.

70. Southall SM, Simpson PJ, Gilbert HJ, Williamson G, Williamson MP. The starch-binding domain from glucoamylase disrupts the structure of starch. FEBS Letters. 1999; 447(1):58–60. PMID: 10218582

71. Kammerer B, Fischer K, Hilpert B, Schubert S, Gutensohn M, Weber A et al. Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter. Plant Cell. 1998; 10(1):105–17. PMID: 9477574

72. Fischer K. The import and export business in plastids: transport processes across the inner envelope membrane. Plant Physiol. 2011; 155(4):1511–9. doi: 10.1104/pp.110.170241 PMID: 21263040

73. Andriotis VME, Pike MJ, Bunnewell S, Hills MJ, Smith AM. The plastidial glucose-6-phosphate/ phosphate antiporter GPT1 is essential for morphogenesis in Arabidopsis embryos. Plant J. 2010; 64(1): 128–39. doi: 10.1111/j.1365-313X.2010.04313.x PMID: 20659277

74. Niewiadomski P, Knappe S, Geimer S, Fischer K, Schulz B, Unte US et al. The Arabidopsis plastidic glucose 6-phosphate/phosphate translocator GPT1 is essential for pollen maturation and embryo sac development. Plant Cell. 2005; 17(3):760–75. PMID: 15722468

75. Kunz HH, Hausler RE, Fettke J, Herbst K, Niewiadomski P, Gierth M et al. The role of plastidial glucose-6-phosphate/phosphate translocators in vegetative tissues of *Arabidopsis thaliana* mutants impaired in starch biosynthesis. Plant Biol. 2010; 12 Suppl 1:115–28. doi: 10.1111/j.1438-8677.2010.00349.x PMID: 20712627

76. Dyson BC, Allwood JW, Feil R, Xu Y, Miller M, Bowsher CG et al. Acclimation of metabolism to light in *Arabidopsis thaliana*: the glucose 6-phosphate/phosphate translocator GPT2 directs metabolic acclimation. Plant Cell Environ. 2015; 38(7):1404–17. doi: 10.1111/pce.12495 PMID: 25474495

77. Bourgis F, Kilaru A, Cao X, Ngando-Ebongue GF, Drira N, Ohlrogge JB et al. Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. Proc Natl Acad Sci U S A. 2011; 108(30):12527–32. doi: 10.1073/pnas.1106502108 PMID: 21709233

78. Ohlrogge JB, Jaworski JG. Regulation of Fatty Acid Synthesis. Annual Review of Plant Physiology and Plant Mol Biol. 1997; 48:109–36.

79. Konishi T, Shinohara K, Yamada K, Sasak Y. Acetyl-CoA carboxylase in higher plants: most plants other than Gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. Plant Cell Physiol. 1996; 37(2):117–22. PMID: 8665091

80. Cernac A, Benning C. *WRINKLED1* encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*. Plant J. 2004; 40(4):575–85. PMID: 15500472

81. Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J. Microarray analysis of developing Arabidopsis seeds. Plant Physiol. 2000; 124(4):1570–81. PMID: 11115875

82. Ke J, Behal RH, Back SL, Nikolau BJ, Wurtele ES, Oliver DJ. The role of pyruvate dehydrogenase and acetyl-coenzyme A synthetase in fatty acid synthesis in developing Arabidopsis seeds. Plant Physiol. 2000; 123(2):497–508. PMID: 10859180

83. Chen M, Mooney BP, Hajduch M, Joshi T, Zhou M, Xu D et al. System analysis of an Arabidopsis mutant altered in de novo fatty acid synthesis reveals diverse changes in seed composition and metabolism. Plant Physiol. 2009; 150(1):27–41. doi: 10.1104/pp.108.134882 PMID: 19279196

84. Hayden DM, Rolletschek H, Borisjuk L, Corwin J, Kliebenstein DJ, Grimberg A et al. Cofactome analyses reveal enhanced flux of carbon into oil for potential biofuel production. Plant J. 2011; 67(6):1018–28. doi: 10.1111/j.1365-313X.2011.04654.x PMID: 21615570

85. Branen JK, Chiou TJ, Engeseth NJ. Overexpression of acyl carrier protein-1 alters fatty acid composition of leaf tissue in Arabidopsis. Plant Physiol. 2001; 127(1):222–9. PMID: 11553750

86. Shockey JM, Fulda MS, Browse JA. Arabidopsis contains nine long-chain acyl-coenzyme a synthetase genes that participate in fatty acid and glycerolipid metabolism. Plant Physiol. 2002; 129 (4):1710–22. PMID: 12177484

87. Tjellstrom H, Strawsine M, Ohlrogge JB. Tracking synthesis and turnover of triacylglycerol in leaves. J Exp Bot. 2015; 66(5):1453–61. doi: 10.1093/jxb/eru500 PMID: 25609824

88. Vanhercke T, El Tahchy A, Shrestha P, Zhou XR, Singh SP, Petrie JR. Synergistic effect of WRI1 and DGAT1 coexpression on triacylglycerol biosynthesis in plants. FEBS Letters. 2013; 587(4):364–9. doi: 10.1016/j.febslet.2012.12.018 PMID: 23313251

89. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science. 2014; 345(6199):950–3. doi: 10.1126/science.1253435 PMID: 25146293

90. Siloto RM, Findlay K, Lopez-Villalobos A, Yeung EC, Nykiforuk CL, Moloney MM. The accumulation of oleosins determines the size of seed oilbodies in *Arabidopsis*. Plant Cell. 2006; 18(8):1961–74. PMID: 16877495

91. Wang X, Devaiah SP, Zhang W, Welti R. Signaling functions of phosphatidic acid. Prog Lipid Res. 2006; 45(3):250–78. PMID: 16574237

92. Lee J, Welti R, Schapaugh WT, Trick HN. Phospholipid and triacylglycerol profiles modified by *PLD* suppression in soybean seed. Plant Biotechnol J. 2011; 9(3):359–72. doi: 10.1111/j.1467-7652.2010.00562.x PMID: 20796246

93. Lu S, Bahn SC, Qu G, Qin H, Hong Y, Xu Q et al. Increased expression of phospholipase Dα1 in guard cells decreases water loss with improved seed production under drought in *Brassica napus*. Plant Biotechnol J. 2013; 11(3):380–9. doi: 10.1111/pbi.12028 PMID: 23279050

94. Reyes JC, Muro-Pastor MI, Florencio FJ. The GATA family of transcription factors in Arabidopsis and rice. Plant Physiol. 2004; 134(4):1718–32. PMID: 15084732

95. Velmurugan N, Sung M, Yim SS, Park MS, Yang JW, Jeong KJ. Systematically programmed adaptive evolution reveals potential role of carbon and nitrogen pathways during lipid accumulation in *Chlamydomonas reinhardtii*. Biotechnol Biofuels. 2014; 7(1):117. doi: 10.1186/s13068-014-0117-7 PMID: 25258645

96. Bi YM, Zhang Y, Signorelli T, Zhao R, Zhu T, Rothstein S. Genetic analysis of Arabidopsis GATA transcription factor gene family reveals a nitrate-inducible member important for chlorophyll synthesis and glucose sensitivity. Plant J. 2005; 44(4):680–92. PMID: 16262716

97. Aubert D, Chevillard M, Dorne A, Arlaud G, Herzog M. Expression patterns of GASA genes in *Arabidopsis thaliana*: the GASA4 gene is up-regulated by gibberellins in meristematic regions. Plant Mol Biol. 1998; 36(6):871–83. PMID: 9520278

98. Chen M, Du X, Zhu Y, Wang Z, Hua S, Li Z et al. Seed fatty acid reducer acts downstream of gibberellin signalling pathway to lower seed fatty acid storage in *Arabidopsis*. Plant Cell Environ. 2012; 35 (12): 2155–69. doi: 10.1111/j.1365-3040.2012.02546.x PMID: 22632271

99. Pinot F, Beisson F. Cytochrome P450 metabolizing fatty acids in plants: characterization and physiological roles. FEBS J. 2011; 278(2):195–205. doi: 10.1111/j.1742-4658.2010.07948.x PMID: 21156024

100. Sun L, Zhu L, Xu L, Yuan D, Min L, Zhang X. Cotton cytochrome P450 CYP82D regulates systemic cell death by modulating the octadecanoid pathway. Nature Commun. 2014; 5:5372.

101. Kim JS, Kim HJ, Lee IA, Lim JS, Seo JY. Regulation of adipocyte differentiation by glyceollins. Faseb J. 2009; 23:712.6.

102. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012; 40:D1178–86. doi: 10.1093/nar/gkr944 PMID: 22110026

103. McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB et al. AgBase: a functional genomics resource for agriculture. BMC Genomics. 2006; 7:229. PMID: 16961921

104. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2007; 23(2):257–8. PMID: 17098774

105. Van Dongen S. Graph Clustering by Flow Simulation [Ph D thesis]. The Netherlands: University of Utrecht; 2000.

106. Aoki K, Ogata Y, Shibata D. Approaches for extracting practical information from gene co-expression networks in plant biology. Plant Cell Physiol. 2007; 48(3):381–90. PMID: 17251202

107. Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD et al. RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol. 2010; 10:160. doi: 10.1186/1471-2229-10-160 PMID: 20687943

108. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods. 2008; 5(7):621–8. doi: 10.1038/nmeth.1226 PMID: 18516045

109. Wang L, Feng Z, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 2010; 26(1):136–8. doi: 10.1093/bioinformatics/btp612 PMID: 19855105

110. Fukuda N, Ikawa Y, Aoyagi T, Kozaki A. Expression of the genes coding for plastidic acetyl-CoA carboxylase subunits is regulated by a location-sensitive transcription factor binding site. Plant Mol Biol. 2013; 82(4–5):473–83. doi: 10.1007/s11103-013-0075-7 PMID: 23733600